

Appendix 1

FILE STRUCTURES

Bibliographic files

The run-time database consists of records each containing a directory, one fixed field and seven variable fields. The directory contains the length of each of the variable fields. The fixed field contains an encoded control number, language, date, Dewey edition number and holdings information. The variable fields are

Author(s) and other names extracted from MARC 100/110/111/700/ 710/711.
Personal names are held as surname and initials. Personal and corporate names are distinguishable.

Main title and subtitle from 245, or occasionally 240 (uniform title).
This is the only field apart from the directory and fixed field which is always non-empty.

Publisher and edition from 260 and 250.

Series and part titles in the following order: 440, 490, 248, 250, 260.
For records with analytical entries any contents note (MARC 505) is included here.

Dewey numbers from MARC 082 or a local field. The first Dewey number is followed by a spine label from a local field (the structure does not cater for different copies of the same work having different shelfmarks).

Sites and number of copies at each site

Subject Headings from 600, 610, 611, 650, 651, 083. (083 in UK MARC contains verbal feature headings derived from PRECIS fields.)

Indexes to bibliographic files

A fairly conventional inverted index system is used. Each index consists of separate primary and secondary indexes, and a postings file.

Secondary indexes are blocked files (block size is four kilobytes), with each block containing a variable number of variable length *index term records*. Each index term record includes a term (which may be a word, a stem, a phrase or a token representing a class of "synonymous" terms), a pointer to the location of this term's postings in the postings file, pointers to the next and previous index term record, and some information about the role(s) of this term in bibliographic records and its number of postings. (When a term has only a few postings these are directly stored in the secondary index instead of in the postings file. This is not of much significance in a subject search system, but saves disk access when searching, for example, a title phrase index, where the great majority of index terms occur in only one bibliographic record.)

A primary index is an indexed sequential file whose records are the last index term in each block of the corresponding secondary index. It is normally held in fast internal memory while a search program is running.

Postings are only 32 bits long. The first 24 bits represent the position

of a bibliographic record, and the remaining eight give some indication of the role or position of the index term in this record. The exact function of the role bits varies from index to index.

An index search proceeds as follows. The sought term is looked up in the primary index using a modified binary search. Usually it will not be found, but what is returned by the primary search is the number of the first primary index record which is greater than or equal to the sought term. Thus if two successive primary index records are "cat" and "dog" and the sought term is "catalogue" the primary search will return the record number of "dog". This will be the number of the secondary index block which will contain "catalogue" if it occurs in the index.

Following the primary search the appropriate secondary index block is read from disk. This is scanned sequentially record by record until either an index term record for "catalogue" is found or an index term greater than the sought term is reached. The procedure returns "success" or "failure" and the address of the index term record which has been found.

This very simple structure does not lend itself to easy updating, but it is quite fast, requiring one disk access and a mean of about a hundred string comparisons for each term search. An otherwise idle machine (Sun 3/50) can look up about 30 terms a second.

Appendix 2

INPUT PROCESSING

This is a more detailed description of the procedures applied to user input outlined in 3.5. Much the same processing is done on bibliographic input during the generation of index terms.

Input is put into lower case and punctuation and spacing are normalized. Repeated blanks are reduced to single blanks. Punctuation not followed by a blank is replaced by a blank (unless it is in a number, or part of an apparent "initialism" like "u.s.a." in which case it is removed). Initialisms containing blanks are also preserved ("u s a" --> "usa"). After this stage the search statement consists only of alphanumerics, blanks, hyphens and apostrophes. the processing of hyphens is fairly elaborate, although they are not very common in user input. Most hyphens are either removed or replaced by a single blank. The sequences " - ", " -" and "- " are all replaced by a blank except when they are apparently part of a range of dates, in which case they are replaced by a single hyphen. Some hyphens are dealt with by look-up (so the arabic "al-" can be removed, and the system also knows some common prefixes like "post", "micro" and "neo"). Two hyphens representing a dash are replaced by a blank. In other cases hyphens are replaced by blanks if there is more than one of them in a "word" or if a single hyphen occupies position five or greater in a word. All other hyphens are removed. Apostrophes are removed unless they are in the second or the penultimate position in a word. Ampersand with a blank on each side becomes " and ". All other characters are removed. If at this stage the search is "empty" the input box clears and the system awaits further input.

The text now consists only of lower case letters, digits, blanks and the occasional hyphen (in date ranges) and apostrophe. In searching, the resulting pre-processed string is what will be shown to the user as the search. Following the pre-processing stage the text undergoes "weak" stemming and spelling standardization. Weak stemming removes regular english plural and "ing" endings with reasonable reliability. Spelling standardization copes with many of the orthographical differences between British and American english - for details see [WALK87b, p65]. The additional "strong" stemming - removing suffixes such as "[iz]ation", "ous" and "ly" - discussed in the earlier report is not used in the current systems, as we concluded that it was rarely beneficial and sometimes detrimental. Strong stemming certainly has a place in searching small or highly specialised databases.

One further point about search processing is worth making. In previous Okapi systems we have always checked searches for duplicate terms ("history of spain and history of portugal" was treated as "history of spain and portugal"). This gave rise to complications in the (admittedly rare) cases when two members of the same synonym class occurred in the same search. When processing "haiti and haitian politics" the system had to display '"haitian" included under "haiti"'. The present systems do not look for duplication. This results in a term which occurs more than once being given (at least) twice the weight it would have if it only appeared once. This only applies to the initial search, not to expanded queries, which never contain repeated terms.

Appendix 3

QUESTIONNAIRES

1 Questions asked before session on dumb system

1. At which college are you a student?

.....

2. Which course are you studying?

.....

3. Which year are you in?

.....

4. Do you use a library in connection with your college work? YES/NO

If YES

5. Which library/ies do you use?

.....

If MORE THAN ONE

6. Which library do you use most frequently?

.....

7. Does this library have a computer for you to use to look up books?

YES/NO

If NO

8. What do you use to look up books?

If YES

9. Do you ever use it? YES/NO

If YES

App 3 Questionnaires

10. How often do you use it?
- a. once a term
 - b. once a month
 - c. once a week
 - d. a few times a week
 - e. every day

11. What sort of books do you look up on the computer?

- a. specific books
- b. books on a topic
- c. both of these

12. Have you ever used a computer in any other library to look up books?

YES/NO

If YES

13. Which library was it?

.....

14. Do you ever use any other type of library catalogue? YES/NO

- a. card
- b. microfiche
- c. printed book
- d. other.....

2 Questions asked after session on dumb system

1. About what proportion of the time did you feel that the computer was useful in helping you to find books which were relevant to the essay titles you were given?

- a. 0%
- b. 25%
- c. 50%
- d. 75%
- e. 100%

Please give reasons for your answer

2. Did you find it helpful to be able to ask the computer to save details of books for you to look at later? YES/NO

Comments/if NO please explain why not

3. Did you have any problems using the computer? YES/NO

If YES

4. Please describe the problems you had.

5. Can you think of any improvements which could be made to the system?

YES/NO

If YES

6. Please describe them

3 Questions asked after session on qe system

1. About what proportion of the time did you feel that the computer was useful in helping you to find books which were relevant to the essay titles you were given?

- a. 0%
- b. 25%
- c. 50%
- d. 75%
- e. 100%

Please give reasons for your answer.

2. Did you find it helpful to be able to ask the computer to save details of books for you to look at later? YES/NO

Comments/if NO please explain why not

3. Did you use the 'MORE' option? YES/NO

If YES

4. Did it help you to find more useful books? YES/NO

Comments/if NO please explain why not

5. Did you have any problems using the computer? YES/NO

If YES

6. Please describe the problems you had.

7. Can you think of any improvements which could be made to the system?

YES/NO

If YES

8. Please describe them.

4 Questions asked after session on full system

1. About what proportion of the time did you feel that the computer was useful in helping you to find books which were relevant to the essay titles you were given?

- a. 0%
- b. 25%
- c. 50%
- d. 75%
- e. 100%

Please give reasons for your answer

2. Did you find it helpful to be able to ask the computer to save details of books for you to look at later? YES/NO

Comments/if NO please explain why

3. Did you use the 'MORE' option? YES/NO

If YES

4. Did it help you to find more useful books? YES/NO

Comments/if NO please explain why not

5. Did you opt to look at books shelved near the one you had chosen? YES/NO

If YES

6. Did this help you to find more useful books? YES/NO

Comments/if NO please explain why

7. Did you have any problems using the computer? YES/NO

If YES

8. Please describe the problems you had.

9. Can you think of any improvements which could be made to the system?

YES/NO

10. Please describe them.

5 Comparison questions asked after subject had used two systems

1. Which system did you find the easiest to use?

- a. First
- b. Second

Why was this?

2. Which system did you find the most helpful in finding books which were relevant to the essay titles you were given?

- a. First
- b. Second

Why was this?

Appendix 4

TASK SHEETS

Art 1

1. Comment on the contribution of art to political propaganda.
2. Did British pop art differ significantly from its American counterpart?
3. What kinds of Gothic were revived in the 19th century?
4. Describe the development of the decorative arts in Britain since 1800.
5. Assess the contribution women have made to art.

Art 2

1. Was it possible for surrealist painting to fulfil surrealist theory?
2. How do the mass media affect popular arts and culture?
3. Discuss the development of parish church architecture in Britain during the middle ages.
4. Discuss painting as a vehicle for social comment through the ages.
5. How important was classical Roman architecture for Brunelleschi?

Computing 1

1. Describe how files are organised under the CP/M operating system.
2. Describe the main components of a digital computer system highlighting the main connections between them.
3. What are the strengths and weaknesses of COBOL compared with other structured programming languages?
4. What factors should be taken into consideration in dialogue design if human-computer interaction is to be problem-free?
5. Describe the role of computers in medicine.

Computing 2

1. Describe a model of a typical microcomputer indicating which levels use compilation and which use interpretation as their method of program translation.
2. Describe some computer architectures which differ from the von Neumann model.

3. How could the installation of a local area network be of benefit to an organisation?
4. What factors should be taken into account when designing a management information system?
5. How can computer data be transmitted through the telephone network?

Engineering 1

1. What is the function of a comparator and what design criteria should it meet?
2. What precautions should be taken when casting metal items to ensure a good quality product?
3. What is the difference between "static pressure" and "stagnation pressure" applied to a fluid and how do they help determine velocity?
4. What equipment is used to manufacture electronic circuit boards?
5. Describe the principles of strengthening processes in metals.

Engineering 2

1. What properties must a cutting tool material exhibit in order to remain effective during the metal-cutting process?
2. What factors should the engineer be aware of to ensure safety in the machine workshop and with what legislation should he comply?
3. Describe the high pressure polymerisation of polyethylene.
4. Describe the Frank-Read mechanism of dislocation generation.
5. When are radioactive isotopes used in engineering?

Life sciences 1

1. Describe the effects insecticides can have on the environment.
2. Compare the relative influence of heredity and culture on intelligence.
3. Discuss the processes by which water, ions etc. move across cell membranes.
4. In what ways has man been seen to adapt to his environment?
5. Outline the mechanisms and principles of plant nutrition.

Life sciences 2

1. Outline the main issues in the biological control of disease

in plants.

2. How accurate are personality tests as predictors of behaviour?
3. Discuss the relationship between host and parasite with specific reference to man.
4. Describe the effects of pollution on the ecosystem of freshwater resources.
5. Discuss the molecular and cellular basis of the immune system.

Social sciences 1

1. Discuss the role of multinational companies in the economics of the third world.
2. Were the Germans right to criticise the peace settlement of Versailles in 1919?
3. Outline the role of social factors in illness.
4. What affects the human's ability to remember?
5. Would the introduction of an electoral system based on proportional representation lead to a more representative government in Britain?

Social sciences 2

1. Discuss the role of the mother in the development of the child.
2. How do sociological perspectives shed light on the role education in society?
3. Would perfectly competitive markets ensure maximisation of social welfare?
4. How widespread was the Slump by 1932? Compare the various national recovery programmes.
5. Assess the significance of the Alliance as a "third force" in British party politics.

General 1

(The general question sheets were used by library and information science students. They are made up from the other question sheets.)

1. When are radioactive isotopes used in engineering?
2. Outline the main issues in the biological control of disease in plants.

3. Discuss painting as a vehicle for social comment through the ages.
4. Discuss the role of the mother in the development of the child.
5. What factors should be taken into account when designing a management information system?

General 2

1. Describe the role of computers in medicine.
2. Outline the role of social factors in illness.
3. What kinds of Gothic were revived in the 19th century?
4. Compare the relative influence of heredity and culture on intelligence.
5. Describe the principles of strengthening processes in metals.

Appendix 5

ANNOTATED EXTRACTS FROM A LOG FILE

Key

Lines from the log are in roman, comments and annotations in *italic*.

The first character in each line identifies the type of data:

- I Identification
- T Date and time to nearest second
- C Comment to improve readability
- S Screen identification
- U User searches and commands
- E Elapsed time to nearest second since last "E" or "T" line.
- L Lookup. Contains enough information to reconstruct what was displayed on the screen.
- O Only used here to show number of items with maximum possible weight, minimum good weight and minimum acceptable weight. These figures can be used to deduce how the result of a search was presented to the user.

The Log

I 22792 - May 25 1988

Log number and date

T 880525095930

Timestamp. Date and time to nearest second.

C 15 f

Subject number and system identification - inserted by the experimenters.

C.....

C ** start

T 880525100135

C ** new topic art1.2

*Question reference (see Appendix *). The question is "Did British pop art differ significantly from its American counterpart?".*

S SCR7.1

Screen is new subject search input screen (Fig 3.).*

U pop aryt british<RET>

Subject enters search, making one correction, and presses Return.

E +62

62 seconds has elapsed since the timestamp.

L pop (10), art (1822), @0227 (12666),

Lookup results. The "token" @0227 represents the class of terms treated as synonymous with "british" ("Britain", "Great Britain", "GB" etc).

E +1

Lookup and merge took a second.

C start

O NMPW = 1 NGW = 10 NAW = 10

One record contains all the terms and there are 10 of "good" weight. The system has retrieved all the records for "pop". Six of them also contain "art". The last four records contain "pop" only, and are all false drops (the last is about a computer programming language).

System reports "1 book matches your search well (10 books found altogether)".

U d

User chooses "Display" option.

S BRIEFS

Brief record display.

U 1

E +41

S FULL

Having spent nearly 41 seconds looking at the brief display, the user chooses to see the first record in full. It is George Melly's "Revolt into style: the pop arts in Britain".

U y

Yes, it is relevant.

U y

"Yes, I would like to see books classified near this one."

E +7

Seven seconds looking at full record and answering the two questions.

S BROWSE_DEWEY at 700.941

S BRIEFS

U d

E +33

User spends 33 seconds looking at the first screen of brief records, then goes forward one screen (nine records) in classified order. Unfortunately 700.941 is a very general number representing art in the British Isles.

S BRIEFS

U d

E +4

S BRIEFS

U d

E +4

S BRIEFS

U b

Having gone forward two more screens the user requests "Back" and returns to the first screen of the records originally found. Reading time for each screen is probably about 3 1/2 seconds (4 seconds less display time).

E +7

S BRIEFS

U 4

E +7

Looks at fourth original record in full ...

S FULL

U y

... and judges it relevant.

U y

E +6

S BROWSE_DEWEY at 709.046

Despite previous experience again chooses to see classified list ...

S BRIEFS

U b

E +14

... but after looking at it for 14 seconds goes back to the original list; although some of the records could well have been useful this is not particularly evident from the one-line displays.

S BRIEFS

U 5

E +10

S FULL

U y

U n

E +14

Chooses another record, but not to see books classified near it.

S BRIEFS
 U m
 E +9
 S MORE
 E +2
 O NMPW = 0 NGW = 0 NAW = 6
Tries the "More" (query expansion) option. The search takes 2 seconds. It finds six records, but they are apparently not very similar to any of the three records chosen so far.
 U d
 S BRIEFS
 U b
 E +33
 S BRIEFS
 U r
 E +4
 S RESTART
 U q
 E +6
 C finish
 T 880525100559
 C ** Finish - duration 264 sec(s)

C ** start
 T 880525100559
 C ** new topic art1.3
"What kinds of Gothic were revived in the 19th century?"
 S SCR7.1
 U 19th century gothic<RET>
 E +34
 L @0013 (738), centuri (2302), gothic (49),
@0013 represents "19", "19th", "XIX" etc. "centuri" is what the stemming procedure makes of "century" or "centuries" (it is not shown to the user like this).
 E +1
 C start
 O NMPW = 2 NGW = 2 NAW = 49
"2 books match your search well (49 books found altogether)"
 U d
 S BRIEFS
 U 1
 E +22
 S FULL
 U y
 U y
Chooses first record and looks at books classified near it.
 E +10
 S BROWSE_DEWEY at 720.942
 S BRIEFS
 U 1
 E +15
 S FULL
 U y
 E +6
 S BRIEFS
 U b
Having chosen one book from classified display returns to original list.
 E +3
 S BRIEFS

U 2

E +6

S FULL

U <RET>y

Here user makes two inappropriate keystrokes. Since neither of these keys is used except during input of a search it is possible that the user was trying to decline to answer the relevance question, an unimplemented feature.

U y

E +7

S BROWSE_DEWEY at 726.50942

S BRIEFS

U b

Has chosen another of the original list, then looked at classified display without choosing any.

E +19

S BRIEFS

U m

E +21

S MORE

E +3

O NMPW = 0 NGW = 3 NAW = 12

'More' finds 12 books, three of them probably matching quite well. The system only reports 'Found some more books, 12 altogether'.

U d

S BRIEFS

U 6

E +34

S FULL

U y

U n

Chooses the sixth book from the query expansion search. It is 'The High Victorian movement in architecture'.

E +5

S BRIEFS

U r

E +6

S RESTART

U q

E +4

C finish

T 880525100915

C ** Finish - duration 196 sec(s)

This was a brisk and decisive searcher who seems to favour short reading lists. He completed nine of the ten questions during his two sessions, a feat equalled by only one other subject and exceeded by none.

Appendix 6

ASSESSORS' INSTRUCTIONS

Instructions

First of all, thank you very much for agreeing to take part in this study.

THE BACKGROUND

Over the past month we have been running experiments to assess the efficiency of a new version of Okapi, the online public access catalogue. The experiments have consisted of a number of students performing subject searches to produce lists of books to assist in answering hypothetical essay questions which we gave them.

We now need the help of people with some reference experience to look through these lists of references and decide to what extent each item would be useful to the person writing the essay.

WHAT TO DO

You have been given one or more lists of bibliographic references. Each list represents the combined results of all the searches conducted on the online catalogue on one topic. The appropriate essay question is printed at the head of each list.

Please read the question carefully.

Then, for each reference:

- a) Use your experience as a librarian to make a quick judgment about the usefulness (or otherwise) of each book to a student writing the essay.
- b) On the scale printed immediately underneath each reference, tick the option which best describes how you feel about the book (we have found that most people take about 10-20 seconds per reference).

When you have finished please return your reference lists to:

.....
.....
.....

Rachel de Vere, Polytechnic of Central London, 74 Gt Portland St, W1.

Example record sent to assessors

Page 1

Outline the mechanisms and principles of plant nutrition

(1)-----[1651571]

AUTHOR(S): SMITH H

TITLE(S): Regulation of enzyme synthesis and activity in higher plants :
proceedings of the Phytochemical Society symposium, Oxford, April
1976.

Annual proceedings of the Phytochemical Society. Symposia series.

PUBLICATION: Academic Press, 1977.

SUBJECT(S): Plant proteins - Congresses. Enzymes - Congresses. Plants.
Biochemistry.

Shelved at : 581.1925 REG

--> Very useful/quite useful/slightly useful/not useful/other--please specify

Appendix 7

CLASS BROWSING RESULTS

This is an expanded version of some of the data summarized in Table 5.6

Question is the topic reference (Appendix 4).

Search statement is what the user typed in.

Number chosen is the number of records chosen relevant by the user from the classified display.

Assessment is the experimenter's opinion of the display (see 5.5.2).

Question	Search statement	Number total	chosen 1st scr	Dewey number	Assessment
=====					
comp1.5	Computers in medicine	0	0	362.172..	remote
		0	0	610.2854..	good
comp1.3	Programming languages cobol	0	0	001.6424	scattered
comp1.4	Human computers [ie human-computer interaction]	0	0	001.64	scattered
comp2.5	computer data in telephone network	2	0	621.38	scattered
		0	0	025.002..	remote
		1	0	001.64044	remote
		0	0	621.38	scattered
comp2.2	computer architectures	0	0	001.64	scattered
		0	0	001.64404	scattered
comp2.4	Management information system	0	0	658.40388	good
comp1.2	Computer architecture and interfaces	1	1	001.64	scattered
		0	0	001.64	scattered
comp1.5	computers in medicine	1	0	610.285..	good
comp1.1	cpm operating system	1	1	001.6425	scattered
art1.5	women art	0	0	709.22	possible
		1	1	759	possible
		0	0	709.22	possible
		0	0	759	possible
		0	0	704.042	possible
		0	0	769.4994..	possible
art1.1	art's contribution to political propaganda	0	0	940.5488	possible
		1	1	303.375	good
soc1.1	multinational companies third world	0	0	769.4994..	possible
		2	1	303.375..	possible
		0	0	909.097..	remote
		0	0	338.891..	possible
soc1.4	memory	7	4	153.12	good
soc1.5	proportional representation	0	0	324.63	good
soc1.5	proportional representation electoral systems	6	3	324.21	possible

App 7 Class browsing results

soc2.2	sociological perspectives education	0	0	370.1934..	possible
		3	3	370.1901	good
		7	0	370.19	good
art1.2	pop art british	0	0	700.941	remote
		0	0	709.046	good
art1.3	19th century gothic	1	1	720.942	good
		0	0	726.50942	possible
art1.1	art politics propaganda	1	1	769.499..	possible
art1.4	decorative arts britain 1800	1	1	747.05	possible
		1	1	747.213	possible
		1	1	745.03	good
art1.5	women art	0	0	759	
soc1.2	treaty of versailles 1919	1	1	943.085	possible
		0	0	943.085	possible
		0	0	943.085	possible
soc1.4	human memory	1	1	616.85232	possible
		17	7	153.12	good
soc1.5	electoral reform	0	0	324.630..	possible
		0	0	324.210..	possible
		1	1	324.25	possible
		0	0	324.21	possible
soc2.1	influence of the mother on the child	2	1	401.9	possible
		1	1	155.422	possible
		0	0	155.422	possible
		1	0	155.418	possible
soc1.5	electoral system pro- portional representation	2	2	324.63	good
	british government				
soc1.4	memory in humans	4	4	153.12	good
		2	2	153.12	good
soc1.3	illness social factors	0	0	616.895	remote
		2	2	306.46	possible
soc1.1	third world multinational companies economic develop- ment	0	0	338.888..	possible
		0	0	338.880..	possible
		0	0	338.880..	possible
soc2.1	role of mother	1	1	155.422	possible
		0	0	155.422	possible
soc2.3	welfare economics	0	0	330.126	possible
		0	0	332	remote
comp1.1	cp/m file	0	0	001.6425	scattered
	operating systems	0	0	001.6425	scattered
comp1.2	computer architecture	1	1	001.64	scattered
life1.3	semi [permeable] membranes	5	2	574.875	good
	ion transport	0	0	541.3724	remote
		1	1	574.875	good
		0	0	581.875	possible
		7	3	574.876	good
eng2.5	engineering and radioactive isotopes	0	0	620.0044	remote
		1	1	541.388	possible
life2.1	biological control and plants and disease	2	2	632.96	good
		0	0	632.96	good
	plants and disease	1	1	632	possible

App 7 Class browsing results

comp2.4	management information system design	1	1	658.403	possible
		0	0	001.64	remote
		0	0	658.403	possible
		0	0	658.403	possible
soc2.1	child developement [sic] and mother	0	0	658.403	possible
		2	1	155.422	possible
		3	1	155.422	possible
		3	3	155.418	possible
comp2.2	computers parallel archi- tecture	3	3	155.418	possible
		0	0	001.64	scattered
		0	0	001.64	scattered
		0	0	001.64044	scattered
comp2.5	computer architecture	0	0	001.64	scattered
		0	0	001.64	scattered
		0	0	001.64	scattered
		0	0	001.64	scattered
soc1.5	lan	1	1	001.64404	possible
		2	2	324.63	good
		2	2	324.63	good
		2	2	324.21	possible
soc1.5	proportional representation britain	1	1	324.630..	possible
		0	0	354.410..	remote
		1	1	324.63	good
		2	2	324.21	possible

Appendix B

EXAMPLE SEARCHES

1 Extracts from a live search on the query expansion system

```
=====
LIST OF BOOKS                                     Books 1 to 9 of 19
Search: "slump 1932"
```

```
-----
No. Title                                     Author          Classmark Date
-----
1 Slump city : the politics of mass unemp.. FRIEND A        331.137941 1981
2 The second slump : a Marxist analysis .. MANDEL E        330.9047   1980
3 The origin and nature of the Great Slum.. FEARON P        338.54094  1979
4 The slump.                                STEVENSON J      941.083    1979
5 The slump : society and politics during.. STEVENSON J      941.083    1977
6 How to survive the slump : a guide to .. BANNOCK G       330.9410.. 1975
7 Politicians and the slump : the Labour .. SKIDELSKY R     941.083    1970
8 Money in boom and slump : an empirical .. WALTERS A A     332.4942   1970
9 Politicians and the slump : the Labour .. SKIDELSKY R     941.083    1967
-----
```

Type its number to see if a book is relevant 3

or type Down (next), Restart/new search/quit

```
=====
Numbers 3, 4 and 5, and possibly others, look relevant. The other 10
books are not relevant, containing only "1932".
```

```
=====
FULL DISPLAY                                     Book 3 of 19
Search: "slump 1932"
```

```
-----
AUTHOR(S): FEARON P
```

TITLE(S): The origin and nature of the Great Slump, 1929-32.
Studies in economic and social history.

PUBLICATION: Macmillan, June 1979.

SUBJECT(S): Depressions - 1929 - Europe. Depressions - 1929 - United States.
Economic depressions, 1929-1938. Related to economic depressions
in United States. Related to economic depressions in Europe.

Not in this branch

No. of copies in other PCL libraries : RHS (2)

Shelved at : 338.54094 FEA 338.540973

```
-----
Is this at all the sort of book you are looking for? (y/n) YES
```

```
=====
This brings in useful terms such as "depression", "economic", United States",
"1929" and "1929-32", as well as unhelpful ones.
```

App 8 Example searches

System has gone back to brief display, and user has chosen "More", which found two books.

```
=====
LIST OF BOOKS similar to the one you chose           Books 1 to 2 of 2
(Original search: "slump 1932")
```

```
-----
No. Title                                     Author          Classmark Date
-----
1  The Great Crash, 1929.                     GALBRAITH J K    338.540973 1980
2  The great crash, 1929.                     GALBRAITH J K    338.540973 1978
  ** END OF LIST **
```

```
-----
Type its number to see if a book is relevant 1
```

Type Back to return to the books you originally found

or type Restart/new search/quit

```
=====
```

```
=====
FULL DISPLAY of books similar to the one you chose       Book 1 of 2
(Original search: "slump 1932")
```

```
-----
AUTHOR(S): GALBRAITH J K
```

TITLE(S): The Great Crash, 1929.

PUBLICATION: 50th anniversary ed. Deutsch, 1980.

SUBJECT(S): Depressions - 1929 - United States. Economic depressions.

Not in this branch

No. of copies in other PCL libraries : RHS (2)

Shelved at : 338.540973 GAL

```
-----
Is this at all the sort of book you are looking for? (y/n) YES
```

```
=====
```

We now have the term "crash" as well.

User has gone back to the original list and chosen another record.

```
=====
FULL DISPLAY                                     Book 5 of 19
Search: 'slump 1932'
```

```
-----
AUTHOR(S): STEVENSON J; (COOK C)
TITLE(S): The slump : society and politics during the depression.
PUBLICATION: Cape, 1977.

SUBJECT(S): Great Britain - History - George V, 1910-1936. Great Britain -
            History - Edward VIII, 1936. Great Britain - History - George VI,
            1936-1952. Great Britain, 1930-1939.
```

```

      No. of copies in this branch : (1)
No. of copies in other PCL libraries : MRd (1) RHS (4)
      Shelved at : 941.083 STE
```

```
-----
Is this at all the sort of book you are looking for? (y/n) YES
```

```
=====

This has brought in some singularly useless LC subject headings, and an
almost useless PRECIS verbal feature heading.
```

User does "More" again, which this time finds five books (based on the three which the user has now selected).

```
=====
LIST OF BOOKS similar to the ones you chose          Books 1 to 5 of 5
(Original search: 'slump 1932')
```

```
-----
No. Title                                     Author          Classmark Date
-----
1  The great crash, 1929.                     GALBRAITH J K   338.540973 1978
2  The great crash, 1929.                     GALBRAITH J K   338.540973 1975
3  America's greatest depression, 1929-1941. CHANDLER L V   338.540973 1970
4  The world economic and social crisis : .. CASTRO F   338.5409.. 1983
5  We're in the money : depression America.. BERGMAN A   791.430973 1972
  ** END OF LIST **
```

```
-----
Type its number to see if a book is relevant
```

Type Back to return to the books you originally found

or type Restart/new search/quit

```
=====

We now have two more editions of Galbraith. One was not previously
retrieved because it had a different subject heading. The book by
Chandler is probably relevant, and the user selects it. The Castro is a
false drop, and the last book is about making films in the depression.
```

The user now has what is probably a reasonable reading list, given the limitations of the database.

```
=====
LIST OF BOOKS YOU HAVE CHOSEN                      Books 1 to 4 of 4
(Original search: "slump 1932")
```

```
-----
No. Title                                     Author          Classmark Date
-----
1** The origin and nature of the Great Slum.. FEARON P        338.54094 1979
2** The Great Crash, 1929.                   GALBRAITH J K   338.540973 1980
3** The slump : society and politics during.. STEVENSON J     941.083 1977
4** America's greatest depression, 1929-1941. CHANDLER L V 338.540973 1970
** END OF LIST **
```

```
-----
Type its number to see further details of one book
```

Type Back to go back to where you were before viewing the 'chosen' list

or type Restart/new search/quit QUIT

```
=====
```

Some books which were not retrieved are yet another edition of the Galbraith with no subject headings at all, and Kindleberger's "The world in depression, 1929-1939". The latter would be retrieved if a somewhat lower cutoff were used.

In fact there are a number of possibly relevant books. Most of them can be retrieved by sifting through the 87 books indexed under "depression" or "slump", but many library databases would be too large for this to be feasible.

2 Notes on a search of the 'full' system

The subject, a student of social sciences at a London college, was tackling the question "How do sociological perspectives throw light on the role of education in society?". She entered "perspective role of education in society". This finds one book matching "well", 135 books altogether. The first book is "The sociology of education: introductory analytical perspectives", with a subject heading "Society - role of education". (It is worth noting that this is the *only* book which would have been found by a system which combined words with an implicit AND.) The "noisy" words "role" and "perspectives" are in some ways helpful here: a search for "education in society" yields many false drops, with the above mentioned book at about position 100 in the sequence, because it is fairly old and items of equal weight display in reverse publication date order.

Dewey classifies "Social aspects of education" at 370.19, and the first book retrieved is classified at 370.193 ("Relations between school and society"). It is interesting that only eight in the first 100 books retrieved by this search are at 370.19. Some clearly relevant books are classified at 373.41 (Secondary education in the United Kingdom). The

user chose the first book, and then looked at books "shelved near". Several on the first screen were clearly relevant, and the subject chose two of them. These brought in the terms "sociology", "sociological" and "educational". Unfortunately, because of the date of publication sequence, they were as old as the initially chosen book. To see more recent works it would have been necessary to go back several screens of 370.193. To see books, equally likely to be relevant, at the more general number 370.19, one would need to go back several more screens. This would have been a more serious problem in a bigger library, but even so, the user, understandably, selected from what was before her eyes. This raises the question of whether it is better to use random ordering in this type of browsing display, or to start the display at the beginning of the sequence for the selected class number. The disadvantage of the latter is that it is impossible to show the "pivot" item in context.

Having chosen two books classified near the one originally chosen, the subject then requested "More". This gives a list of 48 books, mostly at 370.193 or 370.19, with one book at 373.41. She chose number five from the first screen, "The sociology of educational ideas" (indexed by only one of the terms of the original search), and requested books classified near. She chose three books from the first screen and the succeeding one, then went down four screens (she was now back at 370.193) where she chose "An introduction to the sociology of education" and then again requested "More". This gave a list of 95 books, mostly at 370.19 and 370.1. Many of them were, as would be expected, included in the previous query expansion list (of which the subject had only looked at the first screenful and selected one). Near the bottom of this list, and not seen by the user, there were two alarming false drops ("What's left for immunology" and "The management of climatic resources"). These were unfortunately indexed under "inaugural", "delivered" and "February", as was Banks on "Sociology and education", one of the books previously chosen. On the whole, though, "noise" words have remarkably little effect on the results of query expansion searches.

Without looking at the rest of the list, the user chose the first record (Morrish on "Sociology of education"), then looked at records shelved nearby (370.19 again). Before going beyond the first screen she checked on what she had selected using the "View" option - nine records, all with "sociology" and "education" in the title. She then went rapidly forward no less than 48 screens, spending between three and eight seconds looking at each screen. The whole session took about 11 minutes, with nine records chosen, all of which were judged relevant by the assessors.

The 14 subjects who searched for this topic chose a total of 67 distinct titles, and it is not difficult to find more than this. One subject, apparently short of time, chose no records. The other 13 chose an average of 12 books each, with a mean assessed precision of 87% and about 10 brief records seen per choice. The comparative figures for all topics were 6.9 books chosen, mean precision 65% and 9.8 brief records per choice. This is one of the three (out of 50) topics which we rated as easy. Any of the six combinations of a term from "society", "sociology", "sociological" with "education" or "educational" retrieves a list from which one can select some relevant books. These two-term combinations find between 15 and 130 records. As we have seen, the addition of a few woolly qualifiers may not be harmful. It might seem that expansion facilities are not helpful for topics like this one. However, the lists obtained by expanding from one or two good records,

or even the shelf list, are usually very much more "concentrated" than the original lists. An example is the list obtained by the first expansion in the session described above. It retrieved 48 records, of which 16 of the first 18 (two screens) are worth considering. In contrast, the original sequence of records retrieved by "Perspective role of education in society" contains only four possibly-relevant items in the first 18. It is easy to obtain these more concentrated lists, so it is likely that the use of query expansion often reduces the effort involved in even an "easy" search like this one.

It is difficult to see what additional information would help users to choose between one title and another. Our subjects were of course acting more as search intermediaries than as end users. Perhaps an experienced intermediary would have tried to choose a selection of titles which look as though they may treat the subject from different viewpoints - one might choose, for example, texts such as "Schooling and capitalism", "Education and the state", "Family, work and education" as well as some standard-looking books on educational sociology. In contrast, users looking for books to satisfy barely-verbalized personal needs will be influenced in their choices by subjective factors such as knowledge of the author, previous acquaintance, and other, less tangible, aspects, none of which is very useful as feedback information for the system.